J607 Rec'd PCT/PTO 27 DEC 2001

FORM PTO-1390 (REV. 11-2000) U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY 'S DOCKET NUMBER	
42390.P8351	
U S. APPLICATION NO (If known, see 37 CFR 1 5	
10/019882	
INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE
PCT/CN00/00158	15 June 2000 (15.06.00)
PRIORITY DATE CLAIMED	

TITLE OF INVENTION

Speaker Adaptation Using Weighted Feedback

APPLICANT(S) FOR DO/EO/US

YAN, Yonghong

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [ ] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. [ ] The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is attached hereto (required only if not communicated by the International Bureau).
   b. [ ] has been communicated by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. [ ] is attached hereto.
   b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are attached hereto (required only if not communicated by the International Bureau).
   b. [ ] have been communicated by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [X] have not been made and will not be made.
8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [ ] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [ ] A FIRST preliminary amendment.
14. [ ] A SECOND or SUBSEQUENT preliminary amendment.
15. [ ] A substitute specification.
16. [ ] A change of power of attorney and/or address letter.
17. [ ] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. [X] Other items or information:

Return Postcard

I hereby certify that this correspondence is being deposited with the United States Postal Service as Express Mail Label # EL 546 267 088 US in an envelope addressed to: Assistant Commissioner of Patents, US Patent and Trademark Office, Washington, DC 20231, Box PCT.

Greg Kelly — Signature

12/27/01 — Date

531 Rec'd PCT/PTC 27 DEC 2001

U.S. APPLICATION NO (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO	ATTORNEY'S DOCKET NUMBER
10/019882	PCT/CN00/00158	42390.P8351

21. ☒ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):	CALCULATIONS PTO USE ONLY	
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1,040.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO 890.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $740.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) $710.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) $100.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 1,040.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$ 0.00	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	30 - 20 =	10	$18.00	$ 180.00	
Independent claims	4 - 3 =	1	$84.00	$ 84.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			$280.00	$ 0.00	
TOTAL OF ABOVE CALCULATIONS =				$ 1,304.00	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +				$ 0.00	
SUBTOTAL =				$ 1,304.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$ 0.00	
TOTAL NATIONAL FEE =				$ 1,304.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$ 0.00	
TOTAL FEES ENCLOSED =				$ 1,304.00	
				Amount to be refunded:	$
				charged:	$

a. ☒ A check in the amount of $ 1,304.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. 02-2666 in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 02-2666 . A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Mr. John P. Ward
BLAKELY SOKOLOFF TAYLOR & ZAFMANN LLP
12400 Wilshire Boulevard, 7th Floor
Los Angeles, CA 90025-1030
United States of America

SIGNATURE

John P. Ward
NAME

40,216
REGISTRATION NUMBER

4/PRTS 10/019882
531 Rec'd PCT/ 27 DEC 2001

SPEAKER ADAPTATION USING WEIGHTED FEEDBACK

Background

Technical Field of the Invention: The present invention relates to speech recognition systems and, more particularly, to speaker adaptation using feedback.

Background Art: Speech recognition systems using only Speaker Independent (SI) models are very sensitive to different speakers due to speaker characteristic variations. SI models typically use a Hidden Markov Model (HMM). Speaker adaptation is a process to adapt a SI model to a speaker dependent (SD) model to capture the physical characteristics of a given speaker. Speaker adaptation techniques can be used in supervised and unsupervised mode. In supervised mode, the correct transcription is known, while in unsupervised mode, no correct transcription is available.

For reliable and robust speaker adaptation, large amounts of adaptation data are often required in order to cover the linguistic units of a given language. However for most practical applications, only a limited amount of adaptation data is available. Efficient use of the adaptation data becomes extremely important. The traditional adaptation schemes treat all the adaptation data indiscriminately, which results in some parts of the adaptation data being relatively under-trained or under-weighted. Usually the under represented words are more unlikely to be recognized by the decoder.

The traditional adaptation scheme is as follows:

1. Given some adaptation enrollment data and a SI model, collect statistics on the enrollment data and perform speaker adaptation on the SI model.

2. Decoding the test utterances with the adapted acoustic model.

Such a scheme uses the enrollment data only once and does not incorporate any feedback from decoding. It is fast in practice, but does not always provide good performance.

Approaches to speaker adaptation include those described in J. L. Gauvain et al. "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov Chain," IEEE Trans. On Speech and Audio Processing, Vol. 2, pp. 291-298; L.R. Bahl, et al., "A New Algorithm for the estimation of Hidden Markov Model

EL 608 582 035 US

Parameters," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 493-496, 1988; and C.L. Leggetter et al., "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs," Computer Speech and Language, Vol.9, pp. 171-185, 1995. In some of these approaches, errors included in recognizing a particular speaker's utterances are not considered. In a "corrective training" approach, such as in the above-recited L. R. Bahl et al. article, an error in recognition of the utterance may be considered, but a very complicated technique is used to compensate for it. Background articles on expectation maximization (EM) maximum likelihood (ML) are provided in the articles A.P. Dempster, et al., "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal statistical Society, Series B 39, pp. 1-38, 1977; and N. Laird, "The EM algorithm," Handbook of Statistics, vol. 9. Elsevier Science Publishers B.V. 1993.

An iterative technique in speech recognition is to recognize utterances based on an SI model and to create an SD model therefrom and then to apply the SD model to recognizing the utterances to create a more refined SD model and so forth.

There is a need for improved techniques for speaker adaptation. Such improved techniques are described in this disclosure.

Brief Description of the Drawings

The invention will be understood more fully from the detailed description given below and from the accompanying drawings of embodiments of the invention which, however, should not be taken to limit the invention to the specific embodiments described, but are for explanation and understanding only.

FIG. 1 is a partial flow and partial block diagram representation of some embodiments of the invention.

FIG. 2 illustrates a segment (e.g., a phone) of the utterances which includes multiple frames.

FIG. 3 illustrates a section (e.g., word) of the utterances which includes multiple segments (e.g., phones).

FIG. 4 is partial flow and partial block diagram representation similar to a portion of FIG. 1, but may allow multiple feedback passes.

FIG. 5 is a high level schematic block diagram representation of a computer system that may be used in connection with some embodiments of the invention.

FIG. 6 is a high level schematic representation of a hand-held computer system that may be used in connection with some embodiments of the invention.

Detailed Description

The present invention involves speaker adaptation whereby characteristics of an SI model can be adapted through consideration of adaptation enrollment data from a particular speaker to create an SD model. More particularly, the adaptation enrollment data is weighted according to errors detected in the recognized utterances. For those words (or utterances in the enrollment data set) that are not well learnt by speaker adaptation, as indicated by misrecognizing those words, the invention provides a way to incorporate the decoding feedback so that these words can be better adapted. When only limited amounts of enrollment data are available, this scheme of iterative bootstrapping makes better use of that limited data. The scheme can be extended to the unsupervised adaptation where references may contain errors. In some embodiments, an iterative adaptation scheme dynamically adjusts enrollment data to incorporate feedback from decoding on the enrollment data.

In the following disclosure, the term "some embodiments" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the invention. The various appearances "some embodiments" are not necessarily all referring to the same embodiments.
In the following disclosure, when the term phone is used, it could include all phonemes in a particular language or less than all the phonemes. To reduce complexity, some speech recognition systems do not recognize every phoneme in a particular language.

The following four parts are used in some embodiments of the invention. A fifth part is used in still other embodiments.

1. Denote M as the initial SI (speaker independent) model and A as the enrollment data set.

2. Perform speech recognition on data set A based on model M.

3. Adjust A to A' according to the decoding results from part 2. Emphasizing or de-emphasizing certain parts of A with weights based on these results. The emphasizing/de-emphasizing is achieved by assigning a weight to each word in the adaptation data. How to calculate the weight will be discussed below.

4. Adapt model M to M' using enrollment data A'.

5. (optional) Repeat parts 3 and 4 with the updated M'.

For example, FIG. 1 is represents some embodiments of the invention in a diagram which is partially a flow diagram and partially a block diagram. A dashed line represents a dividing line between acts occurring during an adaptation mode and a recognition (decoding) mode. The recognition phase occurs after the SD model is created in the adaptation phase. Note that microphone 14, processing block 18, and recognition block 22 are shown above and below the dashed line and may represent the same blocks at different times (before and after the conclusion of adaptation). In this disclosure, a block may be hardware or a combination of hardware and software.

Referring to FIG. 1 above the dashed line, a speaker input such as microphone 14 receives utterances of a particular speaker. The utterances are converted to digital signals U and may be otherwise processed according to well know techniques by processing block 18. Note that microphone 14 may be adjacent to the computer system that performs the acts illustrated in FIG. 1 or microphone may be remote from it. For example, microphone 14 may be in a telephone or remote other system. Processing block 18 provides the processed utterances U to a recognition block 22 and a weighting block 30. Utterances U may be stored in a wave file as a collection of utterances. Of course, there may be spaces of silence or lack of speech between the sections of the utterances.

Recognition block 22 produces a recognized (hypothesized) phone string H based on the utterances U and an SI model. In a comparison and weight calculating block 26, recognized phone string H is compared with a reference (true) phone string R. The reference phone string is what the speaker is requested to read. A word-phone dictionary may be used to convert the reference word string into phones. Of course, there may be silences or lack of speech in the recognized and reference phone strings.

Differences between the recognized and reference phone string can be determined in a variety of ways. In some embodiments, speech features in the recognized and reference phone strings are compared on a frame by frame level. Merely as an example, the frames may be on the order of about 25 milliseconds (ms), although various other frame durations could be used. A phone may be around 200 milliseconds, although various other phone durations are possible. Accordingly, in some embodiments, there may be on the order of 10 frames per phone. A force alignment algorithm may be used to mark the time or place each phone (or word) happens in the utterances. The frames may contain a Gaussian feature vector.

For example, referring to FIG. 2, portions of the reference string and recognized string for a series of frames are illustrated. The frames are arbitrarily labeled F1, F2, F3, ..., FX-2, FX-1, and FX, wherein there may be several frames between frames F3 and FX-2. The portions in a frame may be a feature extraction. Each portion (e.g., feature extraction) has characteristics (e.g., Gaussian), which are labeled "C." The particular number after the "C" is arbitrarily chosen. For example, in frame F1, both the reference and recognized strings have characteristics C4. Accordingly, the comparison indicates that the characteristics of the reference and recognized strings are the same (S) for frame F1. In frame F2, the characteristic of the reference string is C15 and the characteristic of the recognized string is C11. Accordingly, the comparison indicates that the characteristics of the reference and recognized strings are the different (D) for frame F2. (Merely as an example, S could be "0" and D could be "1," or various other schemes could be used.) Likewise, in frames F3, FX-2, and FX-1, the characteristics are the same and in frame FX, the characteristics are different.

In some embodiments, a certain number of frames forms a segment. The segment may be a phone or other portion of the utterance. Referring to FIG. 2, as an example, a segment 1 may be formed of frames F1, F2, F3, ..., FX-2, FX-1, FX. As illustrated in FIG. 3, a section of the utterances may be formed of multiple segments. In some embodiments, the section is a word, although the invention is not so limited. Segments of silence or lack of speech can be used to indicate the boundary of a word.

If a word includes a phone having an error (the characteristics of a frame of the reference and hypothesis in the word are different (see FIG. 2)), then the word is considered an error word, and the weight of the word is calculated.

The weights assigned to the sections of utterances U may be calculated in block 26 through various techniques. The following are some examples, although the invention is not limited to the examples.

In some embodiments, the weight value for each word is estimated from the likelihood information of the references (the true input word string) and hypotheses (the word string decoded by the recognizer, may contain errors).

1. Run a force alignment program on the reference stream to get statistics of the references.

2. Decode the utterance to get statistics of the 1-best hypothesis.

3. Align the 1-best hypothesis with the reference sentence to obtain the error words.

4. Calculate the average likelihood difference per frame according to the equation (1) as follows:

$$Ln = \frac{H_L^n}{H_e^n - H_b^n} - \frac{R_L^n}{R_e^n - R_b^n} \qquad (1),$$

where H_L^n is the log likelihood of hypothesis word n, H_b^n is the beginning frame index (in time), and H_e^n is the end frame index. R_L^n, R_b^n and R_e^n are the reference counter parts. Of course, the invention is not limited to the details of equation (1). Note that equation (1) involves likelihoods, which are not necessarily probabilities. Equation (1) could be modified to involve probabilities.

Next, the weight value W_i for misrecognized words of a particular speaker "i" is obtained by averaging L_n over all the misrecognized words (error words) according to equation (2) as follows:

$$W_i = \frac{1}{m} * \sum_{n=1}^{m} |Ln| \qquad (2),$$

wherein m may be the number of misrecognized words. Of course, the invention is not

limited to the details of equation (2). In equation (2), the sections are for words (e.g., "W" refers to words), but the sections could be something else. (See FIG. 3.)

Note that in embodiments using equation (2), each misrecognized word may have the same averaged weight. Alternatively, different words could have different weights through applying the result of equation (1) more directly.

Once the weights are calculated, the weights and places of error are provided to block 30. The reference string (or at least the portions of the reference string corresponding to the errors in the recognized stream) is communicated to block 30. In block 30, the utterance U is marked with the errors and corresponding weights are noted. The adaptation enrollment data (E) includes the marked utterances with corresponding weights (wU) and at least those portions of the reference stream (R*) that correspond to the errors in the utterance. The SI model and SD model may be Gaussian mixtures. The wavefile U may be transformed (e.g., through an FFT) from the time domain to the frequency domain. The weight w may be expressed as a floating point number.

In adaptation box 34, the adaptation enrollment data is applied with the SI model to create the SD model according to known techniques, except that the enrollment data may have additional weights. In some embodiments, in the adaptation box 34, the error words are added w times to the SI model. In some embodiments, these weights are added to those of the SI model, although the invention is not limited to this. More complicated schemes could be used, but are not required.

It is important to not give too much weight to the enrollment data, because they are based on limited sampling.

In the above described embodiments, weights are only calculated for words for which there is an error in recognition. Alternatively, there could be weights (e.g., negative weights) for correctly recognized words. Note that in different embodiments the weights can be positive or negative depending on the scheme chosen.

Once the SD model is calculated in an adaptation mode, it is applied on path 40 for use by block 22 in a recognition mode, below the dashed line.

FIG. 4 illustrates that the feedback can be performed more than once until differences between H and R are less than a threshold (see decision block 36). (It could

be less than or equal to a threshold.) To determine whether the differences between H and R are less than a threshold, the various errors can be summed and then compared to a single threshold or different errors can be compared to separate thresholds. Other approaches could be used.

Note that in FIG. 4, the input to recognition block 22 changes with each pass. The utterances may be stored for re-use. The inputs to the adaptation block 24 also changes, however, if the difference between H and R is less than a threshold, then the previous enrollment data is the one applied to path 40 for use during recognition mode (decoding).

There are a variety of computer systems that may be used in training and using a speech recognition system. Merely as an example, FIG. 5 illustrates a highly schematic representation of a computer system 100 which includes a processor 114, memory 116, and input/output and control block 118. There may be a substantially amount of memory in processor 114 and memory 116 may represent both memory that is off the chip of processor 114 or memory that is partially on and partially off the chip of processor 114. (Or memory 116 could be completely on the chip of processor 114). At least some of the input/output and control block 118 could be on the same chip as processor 114, or be on a separate chip. A microphone 126, monitor 130, additional memory 134, and input devices (such as a keyboard and mouse 138), a network connection 142, and speaker(s) 144 may interface with input/output and control block 118. Memory 134 represents a variety of memory such as a hard drive and CD ROM or DVD discs. It is emphasized that the system of FIG. 1 is merely exemplary and the invention is not limited to use with such a computer system. Computer system 100 and other computer systems used to carry out the invention may be in a variety of forms, such as desktop, mainframe, and portable computers.

For example, FIG. 6 illustrates a handheld device 160, with a display 162, which may incorporate some or all the features of FIG. 5. The hand held device may at times interface with another computer system, such as that of FIG. 5. The shapes and relative sizes of the objects in FIG. 5 and 6 are not intended to suggest actual shapes and relative sizes.

Various memories mentioned above (e.g., CD-ROM, flash memory, hard-drive) include computer readable storage mediums on which in instructions may be stored which when executed cause some embodiments of the invention to occur.

If this disclosure states a component, feature, structure, or characteristic "may", "might", or "could" be included, that particular component, feature, structure, or characteristic is not required to be included. If the specification or claim refers to "a" or "an" element, that does not mean there is only one of the element. If the specification or claims refer to "an additional" element, that does not preclude there being more than one of the additional element.

Those skilled in the art having the benefit of this disclosure will appreciate that many other variations from the foregoing description and drawings may be made within the scope of the present invention. Accordingly, it is the following claims including any amendments thereto that define the scope of the invention.

P6053.02

CLAIMS

What is claimed is:

1. A method comprising:

(a) calculating estimated weights for identified errors in recognition of utterances;

(b) marking sections of the utterances as being misrecognized and associating the corresponding estimated weights with these sections of the utterances; and

(c) using the weighted sections of the utterances to convert a speaker independent model to a speaker dependent model.

2. The method of claim 1, wherein parts (a) – (c) are repeated at least once.

3. The method of claim 1, wherein the utterances are converted into a recognized phone string a first time through applying the speaker independent model and thereafter through applying the most recently obtained speaker dependent model.

4. The method of claim 1, wherein the estimated weights are computed through computing an average likelihood difference per frame and then computing a weight value by averaging the average likelihood difference over all the error words.

5. The method of claim 1, wherein average likelihood difference per frame is used to calculate the estimated weights and is computed according to the equation (1) as follows:

$$Ln = \frac{H_L^n}{H_e^n - H_b^n} - \frac{R_L^n}{R_e^n - R_b^n} \qquad (1),$$

where H_L^n is the log likelihood of hypothesis word n, H_b^n is the beginning frame index (in time), and H_e^n is the end frame index, and R_L^n, R_b^n and R_e^n are counter parts for a reference string.

6. The method of claim 5, wherein the weight for misrecognized words of a particular speaker "i" is calculated according to equation (2) as follows:

$$W_i = \frac{1}{m} * \sum_{n=1}^{m} |Ln| \qquad (2),$$

wherein m a number of misrecognized words.

7. The method of claim 1, wherein for a particular speaker, different misrecognized words may have a different weight.

8. A method comprising:

(a) recognizing utterances through converting the utterances into a recognized phone string;

(b) comparing the recognized phone string with a reference phone string;

(c) calculating estimated weights for sections of the utterances;

(d) marking errors in the utterances and providing corresponding estimated weights to form adaptation enrollment data; and

(e) using the adaptation enrollment data to convert a speaker independent model to a speaker dependent model.

9. The method of claim 8, wherein the utterances are converted into the recognized phone string through applying the speaker independent model.

10. The method of claim 8, wherein parts (b) – (e) are repeated until differences between the reference and recognized strings are less than a threshold.

11. The method of claim 8, wherein the utterances are converted into a recognized phone string a first time through applying the speaker independent model and thereafter through applying the most recently obtained speaker dependent model.

12. The method of claim 8, wherein the estimated weights are computed through computing an average likelihood difference per frame and then computing a weight value by averaging the average likelihood difference over all the error words.

13. The method of claim 8, wherein an average likelihood difference per frame is used to calculate the estimated weights and is calculated according to the equation (1) as follows:

$$Ln = \frac{H_L^n}{H_e^n - H_b^n} - \frac{R_L^n}{R_e^n - R_b^n} \qquad (1),$$

where H_L^n is the log likelihood of hypothesis word n, H_b^n is the beginning frame index (in time), and H_e^n is the end frame index, and R_L^n, R_b^n and R_e^n are counter parts for the reference string.

14. The method of claim 13, wherein the weight for misrecognized words of a particular speaker "i" is calculated according to equation (2) as follows:

$$W_i = \frac{1}{m} * \sum_{n=1}^{m} |Ln| \quad (2),$$

wherein m a number of misrecognized words.

15. The method of claim 8, wherein for a particular speaker, different misrecognized words may have a different weight.

16. A memory comprising:

a storage medium having instructions thereon which when executed cause a computer system to perform the following method:

(a) calculating estimated weights for identified errors in recognition of utterances;

(b) marking sections of the utterances as being misrecognized and associating the corresponding estimated weights with these sections of the utterances; and

(c) using the weighted sections of the utterances to convert a speaker independent model to a speaker dependent model.

17. The method of claim 16, wherein parts (a) – (c) are repeated at least once.

18. The method of claim 16, wherein the utterances are converted into a recognized phone string a first time through applying the speaker independent model and thereafter through applying the most recently obtained speaker dependent model.

19. The method of claim 16, wherein the estimated weights are computed through computing an average likelihood difference per frame and then computing a weight value by averaging the average likelihood difference over all the error words.

20. The method of claim 16, wherein average likelihood difference per frame is used to calculate the estimated weights and is computed according to the equation (1) as follows:

$$Ln = \frac{H_L^n}{H_e^n - H_b^n} - \frac{R_L^n}{R_e^n - R_b^n} \quad (1),$$

where H_L^n is the log likelihood of hypothesis word n, H_b^n is the beginning frame index (in time), and H_e^n is the end frame index, and R_L^n, R_b^n and R_e^n are counter parts for a reference string.

21. The method of claim 20, wherein the weight for misrecognized words of a particular speaker "i" is calculated according to equation (2) as follows:

$$W_i = \frac{1}{m} * \sum_{n=1}^{m} |Ln| \qquad (2),$$

wherein m a number of misrecognized words.

22. The method of claim 16, wherein for a particular speaker, different misrecognized words may have a different weight.

23. A memory comprising:

a storage medium having instructions thereon which when executed cause a computer system to perform the following method:

(a) recognizing utterances through converting the utterances into a recognized phone string;

(b) comparing the recognized phone string with a reference phone string;

(c) calculating estimated weights for sections of the utterances;

(d) marking errors in the utterances and providing corresponding estimated weights to form adaptation enrollment data; and

(e) using the adaptation enrollment data to convert a speaker independent model to a speaker dependent model.

24. The method of claim 23, wherein the utterances are converted into the recognized phone string through applying the speaker independent model.

25. The method of claim 23, wherein parts (b) – (e) are repeated until differences between the reference and recognized strings are less than a threshold.

26. The method of claim 23, wherein the utterances are converted into a recognized phone string a first time through applying the speaker independent model and thereafter through applying the most recently obtained speaker dependent model.

27. The method of claim 23, wherein the estimated weights are computed through computing an average likelihood difference per frame and then computing a weight value by averaging the average likelihood difference over all the error words.

28. The method of claim 23, wherein an average likelihood difference per frame is used to calculate the estimated weights and is calculated according to the equation (1) as follows:

$$Ln = \frac{H_L^n}{H_e^n - H_b^n} - \frac{R_L^n}{R_e^n - R_b^n} \qquad (1),$$

where H_L^n is the log likelihood of hypothesis word n, H_b^n is the beginning frame index (in time), and H_e^n is the end frame index, and R_L^n, R_b^n and R_e^n are counter parts for the reference string.

29. The method of claim 28, wherein the weight for misrecognized words of a particular speaker "i" is calculated according to equation (2) as follows:

$$W_i = \frac{1}{m} * \sum_{n=1}^{m} |Ln| \qquad (2),$$

wherein m a number of misrecognized words.

30. The method of claim 23, wherein for a particular speaker, different misrecognized words may have a different weight.

Abstract of the Disclosure

In some embodiments, the invention includes calculating estimated weights for identified errors in recognition of utterances. Sections of the utterances are marked as being misrecognized and the corresponding estimated weights are associated with these sections of the utterances. The weighted sections of the utterances are used to convert a speaker independent model to a speaker dependent model.

P0054 3525 US

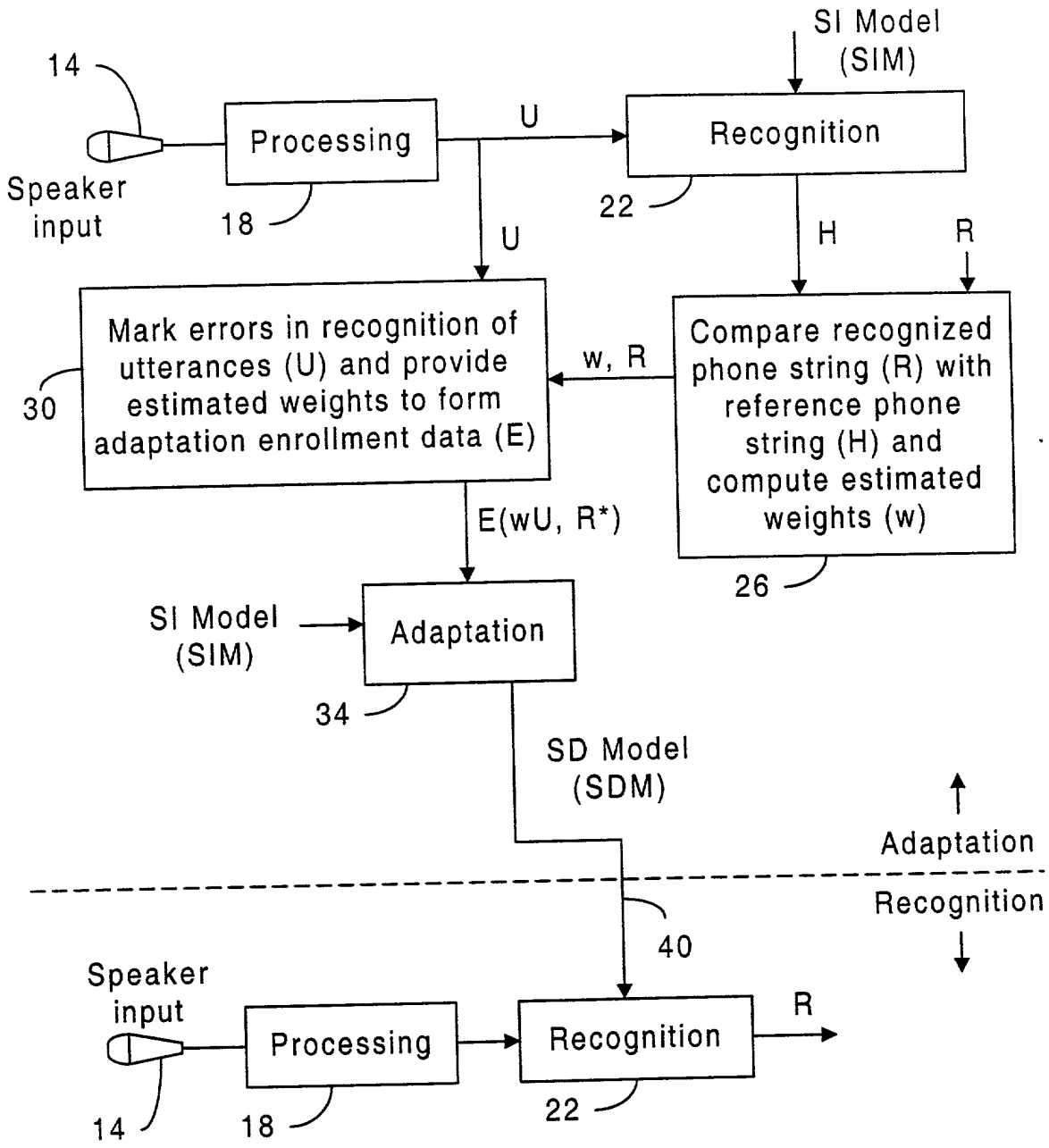
SI Model
(SIM)
14
Processing
U
Recognition
Speaker
input
18
22
U
H
R
Mark errors in recognition of utterances (U) and provide estimated weights to form adaptation enrollment data (E)
w, R
Compare recognized phone string (R) with reference phone string (H) and compute estimated weights (w)
30
E(wU, R*)
26
SI Model
(SIM)
Adaptation
34
SD Model
(SDM)
Adaptation
Recognition
40
Speaker
input
Processing
Recognition
R
14
18
22

FIG. 1

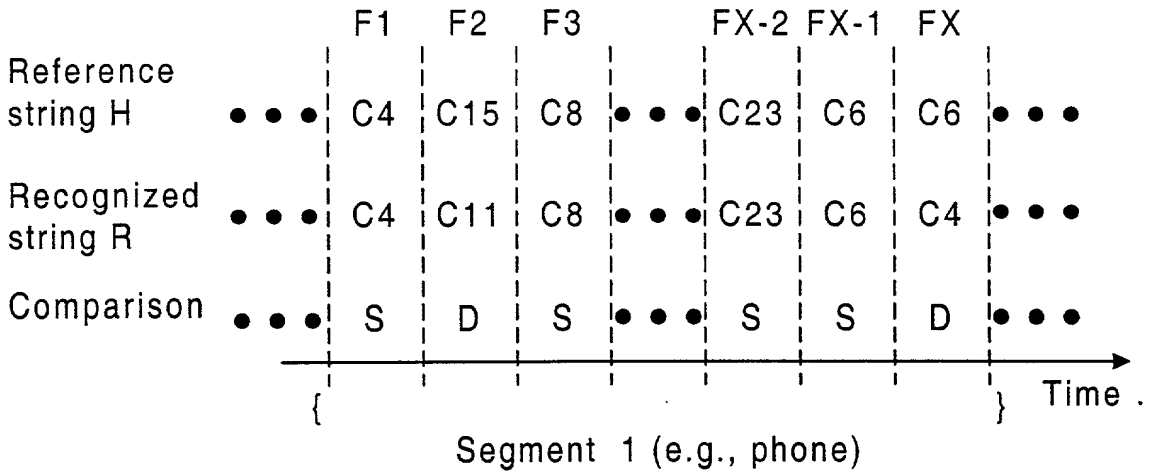

FIG. 2

{ Segment 1 Segment 2 • • • Segment X }

Section (e.g., word)

FIG. 3

Replacement Sheet

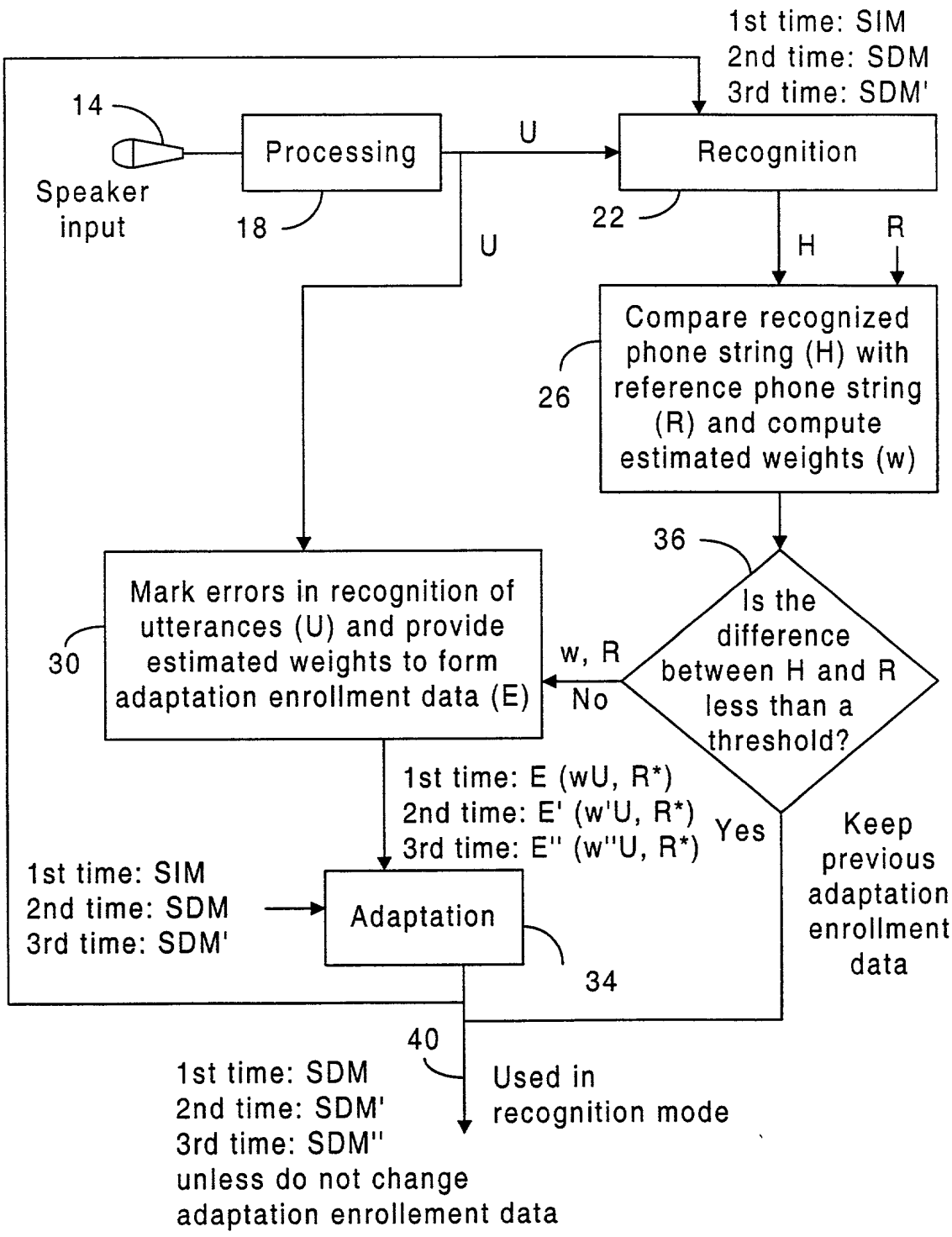

FIG. 4

Docket: 2000P05893US

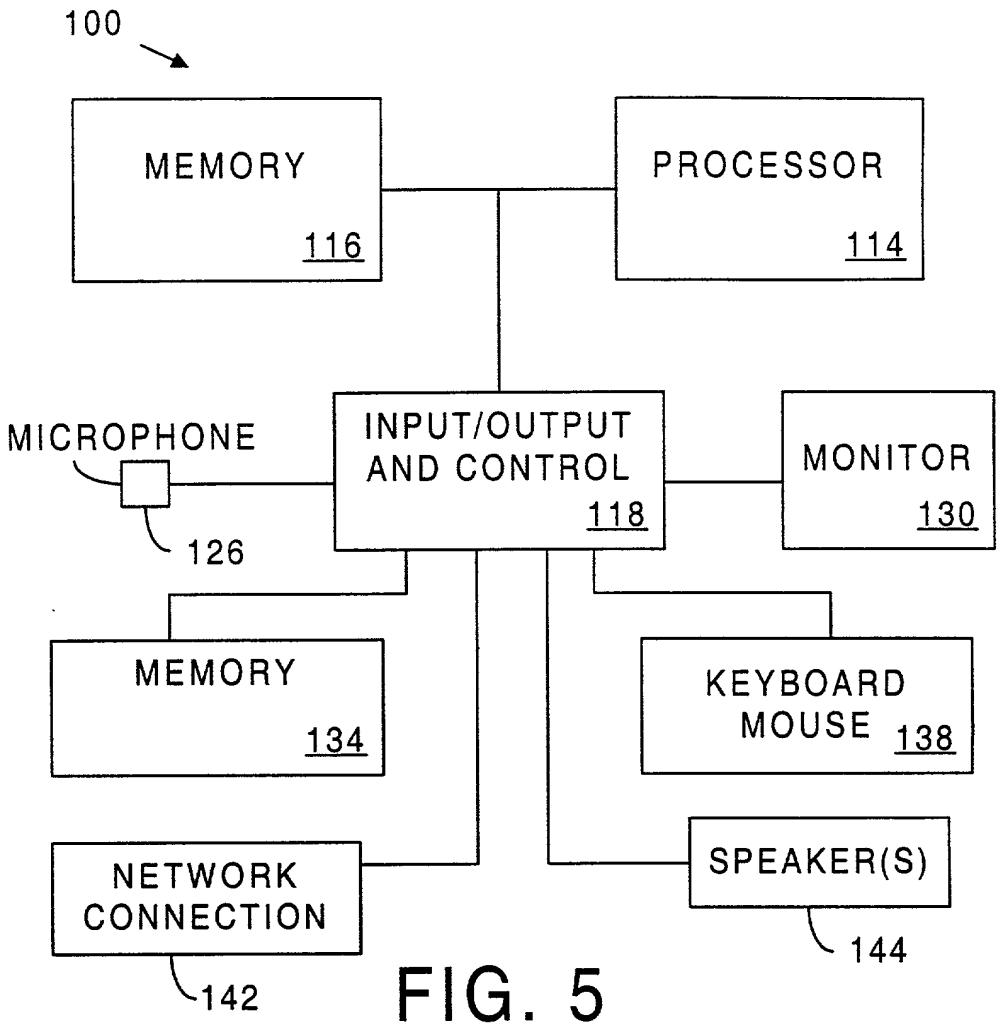
100
MEMORY
116
PROCESSOR
114
MICROPHONE
126
INPUT/OUTPUT
AND CONTROL
118
MONITOR
130
MEMORY
134
KEYBOARD
MOUSE
138
NETWORK
CONNECTION
142
SPEAKER(S)
144

FIG. 5

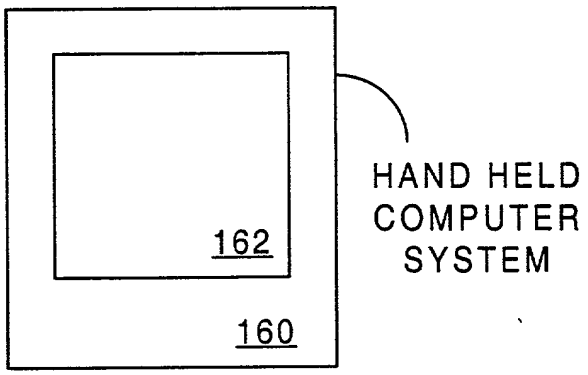
162
160
HAND HELD
COMPUTER
SYSTEM

FIG. 6

Docket No.: **442-010765-US(PAR)**

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Title: **BEARER ADAPTER MANAGEMENT AT A GATEWAY SERVER**

the specification of which

(check one)

☐ is attached hereto.

☒ was filed on as United States Application No.**10/019,892** or PCT International Application Number **PCT/FI00/00512** filed on **7 June 2000** and was amended on (if applicable) **25 July 2001**

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International Application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

(Number)	(Country)	(Day/Month/Year Filed)	Priority Not Claimed
991492	**Finland**	**30 June 1999**	☐
PCT/FI00/00512	**PCT**	**7 June 2000**	☐
			☐
			☐

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

_____ (Application Serial No.) (Filing Date)

_____ (Application Serial No.) (Filing Date)

_____ (Application Serial No.: (Filing Date)

I hereby claim the benefit under 35 U.S.C. Section 120 of any United States application(s), or Section 365(c) of any PCT International Application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International Application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C.F.R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

_____ (Application Serial No.) _____ (Filing Date) (Status) (patented, pending, abandoned)

_____ (Application Serial No.) (Filing Date) (Status) (patented, pending, abandoned)

_____ (Application Serial No.) _____ (Filing Date) (Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

All attorneys listed under Customer No.: 2512

Send Correspondence to:

Customer No.: 2512

Direct Telephone Calls to: (name and telephone number)
Clarence A. Green, Reg. No.: 24,622 (203) 259-1800

Full name of sole or first inventor:
Kyösti RANTO

Sole or first inventor's signature: DATE
Kyösti Ranto 6th February 2002

Residence: YRJÖNKATU 12 K.R. 6.2.02
~~**Suvantokatu 1 D 43,**~~ **FIN-33100 Tampere, Finland** FIX

Citizenship:
Finnish

Post Office Address: K.R. 6.2.02
~~**Suvantokatu 1 D 43,**~~ **FIN-33100 Tampere, Finland**
YRJÖNKATU 12

Full name of second inventor:
Pasi PENTIKÄINEN

Second inventor's signature: DATE
Pasi Pentikäinen 6TH FEBRUARY 2002

Residence address:
Kauppakatu 14 B 12, FIN-33210 Tampere, Finland FIX

Citizenship:
Finnish

Post Office Address:
Kauppakatu 14 B 12, FIN-33210 Tampere, Finland

Full name of third inventor: **Severi EEROLA**	
Third Inventor's signature: Severi Eerola	DATE 7TH FEB 2002
Residence address: **Vuorenhaantie 2, FIN-33960 Pirkkala, Finland** FIX	
Citizenship: **Finnish**	
Post Office Address: **Vuorenhaantie 2, FIN-33960 Pirkkala, Finland**	

3W

Full name of fourth inventor:	
Fourth inventor's signature:	DATE
Residence address:	
Citizenship:	
Post Office Address:	

Full name of fifth inventor:	
Fifth inventor's signature:	DATE
Residence address:	
Citizenship:	
Post Office Address:	

☐ Check here if additional pages are attached. Number of added pages: